

Eliciting stories of gender-transformative change: Investigating the effectiveness of question prompt formulations in qualitative gender assessments

EFFECTIVE QUESTION PROMPTS IN GENDER ASSESSMENTS

Research article

Jess MacArthur, Institute for Sustainable Futures, University of Technology Sydney, Sydney, Australia

Naomi Carrard, Institute for Sustainable Futures, University of Technology Sydney, Sydney, Australia

Veasna Toeur, WASH Department, iDE Cambodia, Phnom Penh, Cambodia

Tyler Kozole, WASH Department, iDE Cambodia, Phnom Penh, Cambodia

Juliet Willetts, Institute for Sustainable Futures, University of Technology Sydney, Sydney, Australia

Corresponding Author: Jess MacArthur

Building 10, 235 Jones St, Ultimo NSW 2007, Australia

jessica.macarthur@uts.edu.au

0000-0002-6645-5767

Declaration of interest: This work was supported by DFAT's Water for Women Fund under Grant WRA-034.

Ms. Jess MacArthur's research explores the social dynamics of water and sanitation programming in South and Southeast Asia and focuses on gender research practices in the development sector.

Professor Juliet Willetts leads applied research to improve development policy and practice and to address social justice and sustainable development, including achievement of the global Sustainable Development Goals (SDGs).

Ms. Naomi Carrard is a Research Director at the Institute for Sustainable Futures. Her applied research focuses on water, sanitation and hygiene (WASH) across the Asia-Pacific region.

Mr. Tyler Kozole is the WASH Program Manager at iDE Cambodia. He leads iDE Cambodia's WASH and climate change programming and research initiatives, focusing on reaching households in climate-vulnerable areas.

Mr. Veasna Toeur is a Monitoring and Evaluation Manager at iDE Cambodia. He leads the monitoring and evaluation team focused on improving WASH systems and gender equality in Cambodia.

Eliciting stories of gender-transformative change: Investigating the effectiveness of question prompt formulations in qualitative gender assessments

Abstract

Evaluations of gender equality initiatives in development programs traditionally assess cognitive dimensions such as knowledge, attitudes, and awareness; and often rely solely on women's perspectives. Leveraging story-based evaluation methods, this article addresses the assessment of complex gender-transformations and focuses on effective question prompts to elicit significant and meaningful narratives of change from both women and men. An assessment process in collaboration with a development program team in Cambodia, led to a set of criteria for considering the quality of respondent stories and testing the efficacy of four different question prompts (n=176): verb-, value-, sphere-, and theme-based. Highlighting aspects of embodiment, the study suggests that verb-based prompts were the most effective at eliciting stories that reflect diverse experiences of both women and men in processes of gender-transformation. Findings from our analysis can support evaluators to balance simplicity and specificity of questions in assessing the unique experiences of individuals undergoing complex change.

Introduction

Over the last decade, efforts to promote gender equality have become increasingly nuanced, focusing on changes within households, workplaces and in the public

sphere (Cornwall and Rivas, 2015). Within this context, programs in the international development sector are aiming to address structural gender inequalities and empower women alongside and within development interventions (Cornwall and Rivas, 2015; Rao and Kelleher, 2005). This trend is mirrored in the private and public sectors spurred on by campaigns for gender parity, equal pay, and ending gender-based violence (World Economic Forum, 2019). Such an approach is referred to as 'gender-transformative', aiming to transform traditional structures, beliefs and practices of masculinity and femininity and fostering an entirely new perspective of gender equality (Cornwall and Rivas, 2015).

For program evaluators investigating gender-transformative change, a core concern is identifying and describing the lived experiences of individuals and how these experiences relate to programs. However, the process of monitoring and evaluating gender outcomes is complex, nuanced, and contextual. There are several reasons for this. First, gendered outcomes associated with interventions can relate to changes in gender equality at individual, relational or structural levels (Carrard et al., 2013; Rowlands, 1995). In addition, the outcomes often intersect with other compounding factors such as race, age, and status, and are different for individuals of different genders (Collins, 2015). This can include aspects such as participation, voice, decision-making, leadership, social norms, and gender roles. Gender equality also must be contextually interpreted (Mohanty, 1984) and it remains difficult to formulate standardised questions that would apply to a range of contexts.

To simplify the evaluation process, many studies speak only to women participants, speak to female and male participants about women's experiences, or focus on pre-selected outcomes that are easily described and understood (MacArthur et al.,

2021). However, in a topic as structural, personal, and nuanced as gender equality, these simplifications have been observed to lead to a fragmented understanding of the embodied experiences of changing gender dynamics (Cornwall and Rivas, 2015).

One approach that addresses these challenges is story-based evaluations, where participants craft their own narratives of change which are shared with program evaluators. This can include longer narrative-based research which uses processes of oral histories and narrative inquiry, and the use of micro-narratives, which capture short descriptions of change (Bamberg, 2006). However, existing examples of story-based gender equality studies remain resource intensive and data quality is inconsistent even with face-to-face facilitation (Bakhache et al., 2017; Bartels et al., 2019; Willetts et al., 2013).

The research presented in this article explores key methodological considerations in designing and using prompts to elicit stories focused on gender dynamics. Prompt design is a critical factor in ensuring that participants both understand and can respond to the line of questioning, and that responses reflect the realms of interest to the assessor or researcher (Smyth, 2016; Yin, 2015). This challenge is particularly pertinent at the time of writing this paper, given limited face-to-face interactions during the COVID-19 pandemic, which has led many evaluators to adopt asynchronous forms of assessment, including self-administered questionnaires.

The article begins with an introduction to the concepts of gender-transformations and embodiment, as critical considerations within studies of personal transformations. Then the paper summarises the components of participatory action research which underpinned this study—a case study conducted in partnership with a civil society

organisation in Cambodia. Each of the four identified question prompts are introduced, followed by a framework to judge the performance of the prompts. Next, the article presents and reflects on the strengths, limitations, and applicability of the four different prompt designs to elicit stories of gender equality. Lastly, the paper discusses both conceptual and methodological considerations for future studies on gender-transformative changes in the international development sector and beyond.

Gender equality as transformation and embodiment

Exploration of methodological issues in eliciting gendered change requires us first to consider what the article means by 'gender equality'. The global goals of gender equality and women's empowerment as articulated in the Beijing Platform for Action (United Nations, 1995) and subsequent Millennium and Sustainable Development Goals (UN General Assembly, 2015) provide a guideline towards a world in which communities are more equal for individuals of different genders. In these goals, gender equality is an outcome of equality and empowerment towards a transformation of society towards "*equal rights, responsibilities and opportunities of women and men and girls and boys*" (UN Women, 2001:1).

Acting on these global mandates, interventions which seek to overcome gender inequalities adopt, often implicitly, one of three equality strategies (Rees, 1998). These strategies shape not only programmatic activities, but also the approaches adopted to monitor and evaluate change. One strategy highlights **sameness**, in which equality is achieved by women entering traditionally male-dominated domains. This conception of gender equality is often assessed through changes in gender parity. A second strategy identifies equality as **difference**, where women and men are equally valued and celebrated for their differences. This is often explored through

changes in the roles that women and men play in society as well as changes in the valuing for non-traditional roles. The third strategy identifies equality as ***transformation***, in which programs aim to transform the structures and norms which govern traditional gender dynamics (Rees, 1998). This third more radical modality is often articulated as a gender-transformative approach (Cornwall and Rivas, 2015; Kabeer, 1994) and new strategies are required to better engage its complexity and nuance.

One lens that has relevance for describing and articulating the contextual nuances of gender-transformative change is that of embodiment. Theories of embodiment claim the importance of a holistic understanding of the person with regards to body, mind, emotions, context, and relationships (Merleau-Ponty, 1962). These aspects resonate strongly with discussions of the ‘gendered self’, in which physical differences form the foundation for structural gender inequalities (Mason, 2018). Much of the extensive literature on gender equality focuses on aspects of cognitive agency such as decision-making, knowledge, control, and autonomy (Ibrahim and Alkire, 2007; Kabeer, 1999; MacArthur et al., 2021). These are all critical aspects, but often do not capture embodied experiences – the ways in which “*bodies reproduce and sometimes challenge gendered power dynamics*” (Mason, 2018: 95).

This paper proposes that by prompting respondents to describe their own embodied experiences, evaluators can begin to untangle the complexity of the cognitive, physical, emotional, contextual, and relational aspects of gender-transformation. The ability of prompts to leverage embodied experience emerged as a finding from this study, as the iterative process of prompt design and analysis generated insights about the performance of prompts against design criteria. As such, the article

introduce ideas of embodiment here to support linear engagement with the article's analysis, noting that embodiment is most relevant to the final tested prompt (verb-based). Findings with reference to embodiment theory are discussed at the end of the paper.

Research Approach

This evaluation research explored both conceptual and methodological challenges of gender equality evaluations by focusing on four unique prompts to investigate gender-transformations for staff of an international development program. To investigate the efficacy of a range of prompts, the research team selected self-administered micronarrative story collection through text and audio submissions. In contrast to traditional forms of narrative inquiry which rely on semi-structured interviews (Adler et al., 2017), micronarratives were deemed an ideal method for prompt exploration as there is no opportunity to probe deeper or further explore responses with participants (Bartels et al., 2019; van der Merwe et al., 2019). The primary research was conducted during the COVID-19 pandemic (July - October 2020), thus reinforcing the value of self-administered approaches.

The research was conducted in collaboration with iDE Cambodia – an international civil society organization with long-term presence in Cambodia. The study was embedded within the SMSU3, operating in six rural provinces, and supporting the promotion and sale of latrines through private enterprises. The program has adopted a gender mainstreaming approach under the Water for Women Fund and therefore our study focused on gendered changes for staff members. The lead author facilitated the research process remotely in collaboration with two research

assistants and coordination support from the program leadership in Cambodia. All field-based staff of the SMSU3 program were invited to participate in the study, and several opted out of participating for a total of 204 participants. The participants of the study all had at least secondary education and had medium to high levels of literacy. This is unique to the documented experiences of gender equality studies using micronarratives which have been more common with individuals with low levels of literacy (Bartels et al., 2019).

The study was approved by the University of Technology Sydney's Human Research Ethics Committee (UTS HREC ETH19-4343), prior to the start of the research. The ethical approval included all components of the study. As gender topics can be sensitive, and domestic violence is prevalent in Cambodia¹, the research team prepared a distress protocol and made connections with relevant feminist organizations for support. All participants were given the opportunity to opt-in to the research and were advised that they were not required to participate, nor would non-participation reflect poorly on them or their work. Participants were informed that the purpose of their contribution was not only to understand their own experiences of the program, but also to help other researchers better ask questions about gendered experiences.

Design thinking-inspired process of data collection

The data collection process consisted of a series of collaborative workshops, reviews, narrative collection, interviews, and observations as summarised in Figure 1 and Table 1. The process was inspired by the three common phases in design thinking (IDEO, 2015): 'hear', 'create' and 'deliver' and adapted to our use-case. This

model is a simplified conception of the actual process, which was inherently iterative and complex.

[Near here: Figure 1]

[Near here: Table 1]

Prompt analysis and assessment

The assessment explored the efficacy of each prompt against the set of criteria identified during the HEAR phase. These criteria will be further explored in a later section of the paper. To assess the prompts through their elicited narratives, micronarrative survey responses were compiled in *Airtable* along with researcher observations and reflections of the survey facilitation. This process enabled each prompt to be reviewed individually and comparatively against each of the four criteria. Analysis was conducted initially by the lead author and then discussed by the research team to ensure coherence and evidence-based insights.

Limitations

This study should be interpreted within its contextual and pragmatic limitations. First, the unique context in Cambodia brings specific gender and social norms. For example, although women often participate in the workforce, there are strong gender codes which govern gender dynamics (Brickell, 2011; Ledgerwood, 1990). The program was also focused on improving sanitation, which historically is also a male-dominated subsector of development programming (Willetts et al., 2010). Second, it should be noted that by directly asking about change, the study pre-supposed that some change had occurred for the staff members. Future iterations of a narrative

survey could include skip logic for participants reporting that they have experienced no change. Lastly, the iterative co-design process and pragmatic objectives of the study led to unequal sampling of each of the four prompt designs. Through a commitment in working in collaboration and creating useful results for the case study program the evaluation prioritised quality stories. Therefore, the sphere-based prompt was not repeated, leading to a lower number of responses. Future studies could use a more even sampling distribution to explore if the patterns identified in this study remain salient.

Design criteria: Dimensions of ‘good’ narratives

In alignment with design-thinking processes, the research began by formulating a set of dimensions which operated both as 1) guidelines to support the design of prompts and 2) as assessment criteria through which to explore the efficacy of participant narratives for generating insights about gender-transformative change. The process can be described as an ‘after-action-review’ in which an intervention is reviewed and refined against its initial objectives. The four criteria, described in Table 2, were derived primarily from the design workshops, which began this research, with inputs and refinements throughout the design and testing processes. Within the workshops, the criteria emerged from the evaluative objectives of the narrative collection: to gather quality and personal stories which described significant changes related to gender equality and associated with the gender mainstreaming intervention. The criteria will be further elaborated within the assessment section of this paper.

Prompt Design: Four testable prompts

Drawing on the systematic review, semantic review, rapid pilot, and pilot reflection

interviews, the paper now introduces each of the four prompts and the design features that emerged from our reviews. For clarity this section includes some methodological aspects to best represent the design process.

The literature review identified a variety of terms used to direct participants towards discussing differences in individual's experiences: changes, effects, impacts, outcomes, experiences, learnings, improvements, and testimonials. The study selected 'changes' as the simplest and most commonly used word within the literature based on learnings from the rapid pilot. This aligns with the Most Significant Change (MSC) approach (Davies and Dart, 2005).

The semantic review explored the phrasing used to draw out stories in both MSC (Davies and Dart, 2005) and Sensemaker (van der Merwe et al., 2019) story elicitation, the two most common narrative approaches in evaluation. Phrasing such as 'share an example', 'talk about your experience', 'describe what happened', 'talk about an experience that illustrates', 'tell a story', 'provide a story' was common in the 23 prompts reviewed. However, the rapid pilot interviews suggested that the words 'story' and 'example', were potentially confusing, with both possibly connoting fictional events within the Cambodian context. Therefore, the study selected wording related to personal experience and 'what happened' in collaboration with Khmer-speaking colleagues.

Building on these insights from the reviews and piloting, the study also investigated best practices related to retrospective self-reporting of changes (Lam and Bengo 2003). This literature suggests that breaking the questions into four parts can increase accuracy and reliability of the stories: 1) what changed? 2) what were things like before? 3) what are things like now? and 4) why did this happen? This

format allows participants to reflect on their own pre-and post-status and then to clarify the direction, extent, and mechanisms of change.

Lastly, after the rapid pilot the study adapted the phrasing of the prompts to clarify the context of change as situated in personal experiences of gender equality. The rapid pilot led to stories which 1) described the current challenges, rather than changes, 2) identified changes un-related to gender equality, and 3) explored changes that were not deemed significant to the respondents. To mitigate these issues, the final assessed prompts included a brief definition of gender equalityⁱⁱ and the phrase “the most significant change for you personally, related to gender equality” to guide participants to discuss their own significant experiences of change about gender equality. This wording is strongly aligned with the Most Significant Change approach (Davies and Dart, 2005).

To provide further context and clarity to each story and to strengthen interpretation in alignment with the intent of the storytellers, the evaluation team also designed a series of short-answer and multiple-choice questions to confirm the meaning and significance of the story and elucidating details such as the types of people involved, and how the story made one feel. These questions were included after the stories sharing section of the micronarrative survey in a similar manner to Sensemaker studies (van der Merwe et al., 2019).

Each of the four prompts are summarised in Table 3 and briefly described below.

[Near here: Table 3]

Value-based prompt - positive and negative stories of change

In the first prompt design, the evaluation directly solicited both positive and negative

stories of change related to the program. This objective was identified during the design workshops with a specific focus on unintended programmatic outcomes. Uncovering unintended outcomes has important implications for the do-no-harm policies which are core to gender equality programs. The word 'value' adopts Jabeen's (2018) phrasing, which identified three value-types of unintended program outcomes: positive, neutral, and negative.

This approach was the third most common modality identified in the literature review (Altenbuchner et al. 2017; Narain 2014; Larson et al. 2018) and specifically aims to identify potential problematic unintended outcomes (Jabeen 2018). Identifying and understanding negative effects is important for program managers of gender equality programming where unintended impacts can lead to domestic violence or backlash (Kabeer 2001). The Most Significant Change literature recommends the inclusion of specific negative change question to ensure that unintended outcomes are identified as *"90 to 95 percent of significant change stories tend to be about positive change"* (Davies and Dart 2005: 19).

For this prompt, participants were first asked about a positive change and then were given the option to share a negative story of change. Emoticons were used alongside the words positive and negative to reinforce intent.

Sphere-based prompt – stories about changes in different spheres of life

In the second prompt design, the study sought narratives related to the locations in which change can take place, with an objective of understanding the extent to which change has impacted other aspects of life outside of the workplace. This effect has been articulated as capability expansion (Keleher, 2014; Sen, 1990) or as the

spillover effect with transformations spilling over from one sphere of life to another (Kabeer, 2005). While this particular approach was not found in the literature review, it has been used by scholars and theorists in articulating the zones in which change can occur (Carrard et al., 2013; Rowlands, 1997).

For this prompt option, participants were shown a set of four digital location cards highlighting changes inside oneself, in the home, in the workplace and in the community. The cards included both Khmer text and a cartoon graphic. Participants were requested to select one card, which represented the sphere in which they had seen the most significant changes related to gender equality and associated with the program.

Theme-based prompts – stories responding to pre-selected themes

In the third prompt design, the evaluation pre-selected themes to guide participants towards specific and known changes. The themes were selected based on commonly associated changes for staff of sanitation programs (Carrard et al., 2013) and in collaboration with program leadership. Identifying themes appropriate for both women and men was a particularly challenging aspect of this prompt.

Based on our systematic review, this prompt design is the most used of the four within literature (see for example: Leahy et al., 2017; Price et al., 2018; Waffi, 2017; Willan et al., 2020). However, in these examples researchers often preselected two or three themes to study, rather than allowing participants to select which themes to discuss.

For this prompt, participants were requested to spend time reviewing sets of digital cards which included both Khmer text and cartoon graphics. Participants were then

asked to select one card to discuss through their micronarrative.

Verb-based prompts – using active words to elicit stories in different domains

Aiming to elicit further stories of personal change, the fourth prompt asked participants to reflect on a set of six verb-based changes and leveraging concepts of embodiment. After the rapid pilot, it became clear that a prompt that simplified language that was accessible for both women and men, could be an interesting and valuable counterpoint to the thematic approach. Drawing on the foundation of embodiment literature, the study identified six verbs to reflect different embodied states relevant to gendered experiences.

These verbs aligned with conceptual domains common in gender equality work. The identification of verbs arose from development of a conceptual framework used in the quantitative analysis of gendered change for WASH programs – the WASH-GEM (Carrard et al., 2022). The WASH-GEM includes five domains which were clarified through verbs for program teams: resources (have and know), agency and structures (do and speak), critical consciousness (think), and wellbeing (feel). The verbs do not capture the full breadth of concepts from the WASH-GEM domains, but offer a simple approach to engage with a set of conceptually complex dimensions.

For this prompt, participants were asked to select one card from a set of six digital verb-based change cards. Each card included Khmer text and a cartoon graphic.

Prompt assessment against the design criteria

Building on the four criteria identified to represent ‘good’ stories (Table 2), this paper now turns to explore how the prompts performed. For clarity, this section includes

brief descriptions of the methodological analysis processes used to assess these aspects.

[Near here: Table 2]

Focused: Narratives related to gender equality

The first assessment criterion highlights the necessity that narratives actually describe changes in gender equality in alignment with the evaluation questions and purpose. During the rapid piloting, the question prompts did not directly include the phrase 'gender equality', to avoid prescriptive or normative conceptualizations and instead relied on gender-related themes, such as respect and confidence. However, this led to confusion, for example responses about respecting authority and elders. Therefore, the tested prompts all included brief descriptions of gender equality (pre-prompt) to ensure that narratives collected were focused to the main premise of the evaluation. Once steps were taken to include a brief description of gender equality within the prompts, all prompts were successful in collecting gender-equality focused narratives. Excluding the blank and non-coherent stories, the remaining stories were focused on narratives of gender equality.

Although gender equality remains a highly contextualized concept, our analysis suggests that overall research participants had similar conceptual models of equality to one another and overall saw changes towards equality as positive. However, this could be because the majority of the collected stories were positive, despite specifically seeking negative stories. The evaluation team surmised that the relatively similar mental models of gender equality within our sample was directly related to the impact of the program. All staff members had participated in the same

gender mainstreaming training, which discussed concepts of gender equality in detail and share a particular normative view of gender equality.

Personal: Narratives of experienced change

The extent to which narratives were connected to personal experiences rather than perceived or observed experiences (non-personal) is highlighted in Figure 2. The evaluation aimed to collect first-person stories to improve trustworthiness and validity in the stories. The proximity of the storyteller to the story was identified by coding the verbs within each narrative. Where applicable, this was informed by each respondent's response to multiple choice questions about who was involved in the story. Verbs of observation are illustrated in light grey, while verbs of personal experience are in darker colours. An example of a non-personal story is included in Box 1.

The verb-based approach performed the best in capturing personal stories of change (90%), while the sphere-based performed the worst (only 20% of the stories were personal). Additionally, as illustrated in Figure 2, the types of stories elicited were explored based on the main verb used within the story. The verb-based prompt, in alignment with its main design objective, was most effective at eliciting a breadth of experiential stories.

[Near here: Figure 2]

[Near here: Box 1]

Programmatic: Narratives associated with program interventions

The second criterion explores the strength of the connection of the narrative to the

gender mainstreaming intervention. The evaluation aimed to elicit narratives of change which were influenced by the intervention. The extent to which the program was perceived to influence change, was coded by the researcher focusing on the sub-question '*why did this change happen?*', which was included in all four prompt designs. If a respondent described both general societal changes and program activities, this was coded as a medium connection. An example of a non-programmatic story is included in Box 2.

As can be seen in Figure 3, the verb-based prompt was the most successful in collecting programmatic stories, while the theme was the least successful. The sphere-based prompts were as effective as the verb-based prompts in eliciting stories somewhat or strongly connected to the program.

[Near here: Figure 3]

[Near here: Box 2]

Coherent: Narratives that include enough detail to analyse

The third area of assessment reviews the coherence of the story from an evaluative perspective. As with any evaluation, the quality of the data elicited from respondents is directly connected with the ability of an evaluator to conduct an assessment. Story coherence was coded based on the presence of five aspects commonly associated with narrative storytelling: who, what, where, when, and why. The presence of an aspect is illustrated with a zero-to-five-star coding process, with one star awarded for each described aspect. Examples of stories with one, three, and five stars are included in Box 3. These scores are often, but not always, related to the length of the story; stories were an average of 142 words long (after transcription and translation

into English).

As illustrated in Figure 4, the verb-based prompt also performed the best at collecting stories that incorporate all five aspects. However, when comparing four- and five-star stories, all but the theme-based prompt performed similarly.

[Near here: Figure 4]

[Near here: Box 3]

Reflections on the efficacy of each prompt

The study will now reflect on the efficacy of each prompt, synthesising the results and reflections with regards to each prompts' particular objective.

Value-based prompt

Although the value-based prompt was successful in eliciting personal stories, it did not effectively elicit negative stories of change – which was its main objective. Most respondents did not share a negative story and instead used the opportunity to share a second positive story, or to reiterate their first story. Value was also explored for all four prompts in the survey section just after the story-sharing, by asking participants to respond to two multiple choice questions.^{iii,iv} Ninety-seven percent (155/169) of total responses elicited feelings of 'happy' and 'very happy', with only one response eliciting 'unhappy' and one 'sad'. Drawing on researcher coding of the value of the stories, the theme-based prompt was the most effective at exploring negative changes.

While in this pilot, the value-based prompt and the whole study more broadly, were not successful at eliciting unintended outcomes, the results highlight the difficulty in

eliciting negative stories, a problem recognized in our semantic review (Davies and Dart, 2005). However, this finding is in contrast to other with studies from our literature review which did not have the same challenges in soliciting negative stories of change. These studies used prompts such as “tell me about the negative effects of organic farming on women” (Altenbuchner et al., 2017) or asking about personal wellbeing in post-conflict situations and following up with “how does this story make you feel?” (Roupetz et al., 2020). As these example prompts were very similar to our prompt designs and the inclusion of a bespoke prompt to collect negative impacts, these results suggests that the context may have hindered sharing of negative impacts. The studies that more effectively elicited negative stories took place in India (Altenbuchner et al., 2017), Pakistan (Asghar et al., 2018) and with Syrian refugees in Lebanon (Roupetz et al., 2020).

Upon further reflection, the team has explored if the “honour and shame” culture in Cambodia precludes the sharing of negative stories. In such a culture, individuals are not prone to share negative reflections without significant probing and rapport-building (Ledgerwood, 1990). This is even more pronounced in internal evaluations, where respondents, even anonymously, are less open to sharing challenges. Nonetheless, a lack of negative stories does not imply that negative experiences are not occurring.

Sphere-based prompt

In a similar manner to the counter-intuitive result of the value-based prompts, sphere-based prompts were less effective at eliciting a diversity of story locations. Additionally, the inclusion of the 'self' as a sphere of change was perhaps misunderstood as it did not elicit the stories around cognitive or personal changes as

intended within the survey. Each prompt also included a multiple-choice question, to further understand the location of change: “who else was involved in this change?”, which included location-specific wording around work, community, and family. Uniquely, the other three prompts were more effective at identifying changes that have spread to home and the community, yet *all* the sphere-based responses focused on work-related change.

As the sphere-based prompt was not as successful at exploring the expansion of change into other realms as the other prompts, our results suggest that using physical locations as prompts may lead participants to reflect on the activities they observe within locations and not on their own experiences in the physical space. Nonetheless, much has been written about the value of exploring gendered space (McDowell, 1999) and the gender divisions of spheres (Rosaldo et al., 1974). The cultural, social, and ultimately structural significance of gendered space reinforces the importance of investigating change across multiple arenas – either directly as a prompt or through follow-up questions.

Theme-based prompt

While the responses to the theme-based prompts were less personal, they provided value in expanding and enriching the program’s understanding of foreseen (intended and unintended) outcomes. The theme-based prompts led respondents to focus more on changes they were seeing in colleagues and family members, rather than on their own experiences of change. The evaluation team believe this is because the visuals generated unique connections with participants’ day to day observations of change. This prompt also included the largest cluster of negative stories. To further understand the breadth of changes that all four prompts elicited, themes were coded

to the other three prompt responses, by the research team. However, this analysis process was complex, with themes overlapping and intersecting in almost all stories of change. The analysis led to the development of personas to represent the types of changes rather than traditional thematic analysis.

Our study revealed the tension between breadth and depth within thematic investigations. Examples in the literature rely on a smaller set (between one and five) of pre-selected themes within the context of in-depth interviews or focus group discussions and based on programmatic theories of change. It was also difficult to ensure the same themes for women and men, as wording became complicated. The theme-based prompt was also unable to capture nuances such as *'changes in confidence to speak with women'* in contrast to the card which read *'changes in confidence for women'*. However, the theme-based prompt was effective at eliciting stories, which were simpler to cluster and analyse, in contrast to conducting thematic analysis with the other three prompts. The use of micronarratives allowed for a wider number of participants and the inclusion of a larger number of themes, but this may not always be practical.

Verb-based prompt

The responses to the verb-based prompt reflected a wide range of personal changes across all spheres and types of change, highlighting the effectiveness of verbs in soliciting diverse narratives as well as embodied change. A low number of respondents (38%) believed that others were also experiencing their change (the lowest of all four prompts), which suggests that this prompt was able to capture a much more diverse and nuanced breadth of responses. The verb-based prompt was also the most successful in capturing changes within the home. This prompt was,

however, less effective at identifying unintended outcomes, indicating an opportunity for future research. Comparisons of the verb-based prompt's objective to collect a range of personal experiences was conducted by coding the verbs that occurred within each story. As illustrated previously in Figure 2, the verb-based prompt was the most effective, despite being unable to elicit a story about a change in what 'you have'.

The simplicity yet specificity of this verb-based prompt, led to a breadth of responses. However, such diversity required more time and resources for data analysis as noted above. Little has been written about the value of verb-based prompts, outside best practice recommendations for qualitative interviews in keeping prompts short, concise, and free from academic jargon. For example, Patton (2002) describes a prompt design approach which reviews the experiences, opinions, feelings, knowledge, and sensations of participants in the past, present, or projected into the future (Patton, 2002: 351–352). Additionally, the verb-based prompt also loosely align with aspects of Empathy Mapping, an interview and analysis tool used frequently in user experience design research (Gray, 2013). Our case study suggests that verb-based prompts are effective, because they are easily accessible to participants and reduce the number of possible response iterations.

Reflections on embodiment and future research

Embodiment as a lens to interpret gender-transformations

The article now considers the usefulness of embodiment as a lens through which to explore gender-transformations. Our interest in embodiment arose from reflections on why the verb-based prompts performed significantly better than the other three

tested prompts. Not only did it collect results that more closely aligned with the prompt intent and evaluation purpose, but the verb-based prompts were most effective in eliciting personal, programmatic, and coherent stories of change. The analysis identified two reasons why embodiment is a useful lens for qualitative assessments of gender-transformations: first, that embodiment is not just women-focused, and second, that it was easiest for participants to understand and respond to the questions.

The evaluation of gender-transformations requires a tool to elicit stories which captures each individual's engagement with change. Embodiment centres on change within the gendered person, recognizing that not only women, but also men, experience change in a transforming environment (Mason 2018). This highlights the ability and responsibility for each individual to become actors of change through processes of critical consciousness (Freire 1970); broadening the narrative of women's empowerment which dominates thinking in the gender equality sector (Cornwall and Rivas 2015). For all prompts, except for the verb-based example, participants discussed the actions and behaviours of others (often women), rather than their own experiences.

The evaluation of gender-transformations requires a tool to simply direct participants to engage with a breadth of possible outcomes. Embodiment, and in particular our usage of embodiment as action verbs, simplifies a range of experiences into a set of six easily recognizable activities. These actions align with the physical, cognitive, emotional, and relational components of embodiment theory (Merleau-Ponty 1962); while also aligning with aspects of gender equality including resources, agency, structures, and wellbeing. Yet this approach circumvents the academic language,

which is not appropriate in many research environments. The sphere- and value-based prompts did not elicit a breadth of outcomes, and the theme-based prompt required more time for respondents to digest the content and select a card, with implications to respondent fatigue and need for additional time and resources.

Implications for evaluators

Evaluators exploring themes of gender-transformation and in particular changes for both women and men are faced with the challenge of asking good questions. Our study has proposed that ‘good’ questions guide respondents to discuss both foreseen and unforeseen outcomes and are not prescriptive. Our study suggests that evaluators can use verb-based question prompts to elicit rich and personal stories of change with program beneficiaries and staff. This approach may also be valuable for other complex topics such as resilience to climate change and market systems change. Additionally, the study has highlighted the usefulness of visual cues through cards and emoticons can help guide respondents to discuss the topic of their choice and open more opportunities to identify unintended outcomes.

The extent to which gender-transformations were discussed by participants may have been influenced by the shared definitions of gender equality between participants as a result of gender training. For evaluations exploring gender-related change in programs where this has not occurred, further prompt testing and co-design would be required to ensure that the prompts are not inadvertently normalising equality and therefore limiting the evaluation insights.

Directions for future scholarship

Few studies have explored how prompts and question formulations impact

qualitative studies and further research could beneficially explore the salience of our findings in other contexts. For example, future studies could explore wider geographic diversity and the use of a range of prompts in different types of gender mainstreaming programming. Studies could also leverage other methods such as interviews or focus groups, which could include cognitive testing of the prompts. Studies could also expand on ways to use verb-based prompts to elicit stories related to unintended and potentially negative outcomes. Additionally, studies could ask about changes in each sphere, instead of asking participants to select a sphere where the most change had occurred. Lastly, prompts that focus on learnings or differences for different types of people, which were identified in our literature review (Dutt and Grabe 2019; Asghar et al. 2018) but not included in our study, could offer a wider breath of possible prompt designs for gender practitioners focused on researching and evaluating gendered change.

Conclusions

As many evaluation practices move online to self-administered and remote modalities, simple and reliable methods for qualitatively exploring complex changes are becoming even more important. Yet studies into the semantic design of question prompts for evaluations are nascent. This paper has described a systematic and collaborative study concerning the design and assessment of four unique question prompts used in evaluating complex gender-transformations for staff of a development program in Cambodia.

Through an assessment of verb-, value-, sphere-, and theme-based prompt designs (n=176), our study shows that the verb-based prompt was the most effective at

eliciting personal, coherent, programmatic, and focused stories of gender-related change. Our analysis suggests that the verb-based prompt was the most successful because it leveraged concepts of embodiment in which all participants (both women and men) were able to respond directly from their own perspective about their own experiences. This is unique, especially in gender equality assessments, which often focus on women rather than the gendered experiences of all individuals.

Findings from this analysis can inform not only gender-focused evaluations, but also studies of other complex processes of change in aspects such as systems strengthening and resilience building.

ⁱ 18.2% of women have experienced physical or sexual violence committed by husband/partner (National Institute of Statistics [Cambodia], 2014)

ⁱⁱ “Both this study and SMSU3 are focused on gender equality how experiences and interactions are becoming more equal between women and men”

ⁱⁱⁱ Question 1: **Pick and share an emoji that describes how you feel about this change.** Select one (1) or add your own option. (*very happy, happy, unsure, sad, unhappy, mad, other*). Emoticons (emojis) were identified during our design workshops as a useful tool due to the ubiquitous nature of visual cues within social media in Southeast Asia.

^{iv} Question 2: **What is the outcome of this change?** Click one (1) option (*positive, neutral, negative*)

References

- Adler, J. M., Dunlop, W. L., Fivush, R., Lilgendahl, J. P., Lodi-Smith, J., McAdams, D. P., McLean, K. C., Pasupathi, M., & Syed, M. (2017). Research Methods for Studying Narrative Identity: A Primer. *Social Psychological and Personality Science*, 8(5), 519–527. <https://doi.org/10.1177/1948550617698202>
- Altenbuchner, C., Vogel, S., & Larcher, M. (2017). Effects of organic farming on the empowerment of women: A case study on the perception of female farmers in Odisha, India. *Women's Studies International Forum*, 64, 28–33. <https://doi.org/10.1016/j.wsif.2017.09.001>
- Asghar, K., Mayevskaya, Y., Sommer, M., Razzaque, A., Laird, B., Khan, Y., Qureshi, S., Falb, K., & Stark, L. (2018). Promoting Adolescent Girls' Well-Being in Pakistan: a Mixed-Methods Study of Change Over Time, Feasibility, and Acceptability, of the COMPASS Program. *Prevention Science : The Official Journal of the Society for Prevention Research*, 19(8), 1030–1042. <https://doi.org/http://dx.doi.org/10.1007/s11121-018-0890-9>
- Bakhache, N., Michael, S., Roupetz, S., Garbern, S., Bergquist, H., Davison, C., & Bartels, S. (2017). Implementation of a SenseMaker® research project among Syrian refugees in Lebanon. *Global Health Action*, 10(1). <https://doi.org/10.1080/16549716.2017.1362792>
- Bamberg, M. (2006). Stories: Big or small: Why do we care? *Narrative Inquiry*, 16(1), 139–147. <https://doi.org/10.1075/ni.16.1.18bam>
- Bartels, S. A., Michael, S., Vahedi, L., Collier, A., Kelly, J., Davison, C., Scott, J., Parmar, P., & Geara, P. (2019). SenseMaker® as a monitoring and evaluation

tool to provide new insights on gender-based violence programs and services in Lebanon. *Evaluation and Program Planning*, 77(June), 101715.

<https://doi.org/10.1016/j.evalprogplan.2019.101715>

Brickell, K. (2011). The “stubborn stain” on development: Gendered meanings of housework (non-)participation in Cambodia. *Journal of Development Studies*, 47(9), 1353–1370. <https://doi.org/10.1080/00220388.2010.527955>

Carrard, N., Crawford, J., Halcrow, G., Rowland, C., & Willetts, J. (2013). A framework for exploring gender equality outcomes from WASH programmes. *Waterlines*, 32(4), 315–333. <https://doi.org/10.3362/1756-3488.2013.033>

Carrard, N., Macarthur, J., Leahy, C., Soeters, S., & Willetts, J. (2022). The water, sanitation and hygiene gender equality measure (WASH-GEM): Conceptual foundations and domains of change. *Women’s Studies International Forum*, 91(2022), 102563. <https://doi.org/10.1016/j.wsif.2022.102563>

Collins, P. H. (2015). Intersectionality’s Definitional Dilemmas. *Annual Review of Sociology*, 41, 1–20. <https://doi.org/10.1146/annurev-soc-073014-112142>

Cornwall, A., & Rivas, A. M. (2015). From ‘gender equality and ‘women’s empowerment’ to global justice: reclaiming a transformative agenda for gender and development. *Third World Quarterly*, 36(2), 396–415. <https://doi.org/10.1080/01436597.2015.1013341>

Davies, R., & Dart, J. (2005). The “Most Significant Change” (MSC) Technique: A Guide to Its Use. <https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf>

Dutt, A., & Grabe, S. (2019). Understanding Processes of Transformative Change: A

Qualitative Inquiry into Empowering Sources and Outcomes Identified by Women in Rural Nicaragua. *Sex Roles*, 81(7–8), 487–504.

<https://doi.org/10.1007/s11199-019-1005-1>

Freire, P. (1970). *Pedagogy of the oppressed*. Continuum.

Gibbs, A., Corboz, J., Chirwa, E., Mann, C., Karim, F., Shafiq, M., Mecagni, A., Maxwell-Jones, C., Noble, E., & Jewkes, R. (2020). The impacts of combined social and economic empowerment training on intimate partner violence, depression, gender norms and livelihoods among women: An individually randomised controlled trial and qualitative study in Afghanistan. *BMJ Global Health*, 5(3), 1–17. <https://doi.org/10.1136/bmjgh-2019-001946>

Gray, D. (2013). *Gamestorming*. O'Reilly Verlag.

Ibrahim, S., & Alkire, S. (2007). Agency and empowerment: A proposal for internationally comparable indicators. *Oxford Development Studies*, 35(4), 379–403. <https://doi.org/10.1080/13600810701701897>

IDEO. (2015). *The Field Guide to Human-Centered Design*. Design-Kit.

Jabeen, S. (2018). Unintended outcomes evaluation approach: A plausible way to evaluate unintended outcomes of social development programmes. *Evaluation and Program Planning*, 68, 262–274.

<https://doi.org/10.1016/J.EVALPROGPLAN.2017.09.005>

Kabeer, N. (1994). *Reversed realities: gender hierarchies in development thought*. Verso.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, 30(3), 435–

464. <https://doi.org/10.1111/1467-7660.00125>

Kabeer, N. (2001). Conflicts over credit: Re-evaluating the empowerment potential of loans to women in rural Bangladesh. *World Development*, 29(1), 63–84.

[https://doi.org/10.1016/S0305-750X\(00\)00081-4](https://doi.org/10.1016/S0305-750X(00)00081-4)

Kabeer, N. (2005). Gender equality and women's empowerment: A critical analysis of the third millennium development goal. *Gender and Development*, 13(1), 13–24. <https://doi.org/10.1080/13552070512331332273>

Mohanty, C. T. (1984). Under Western Eyes: Feminist Scholarship and Colonial Discourses. *Boundary 2*, 12(3), 333–358.

Keleher, L. (2014). Sen and Nussbaum: Agency and Capability-Expansion. In *Éthique et économique/Ethics and Economics* (Vol. 11, Issue 2). <http://ethique-economique.net/>

Lam, T. C. M., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24(1), 65–80. [https://doi.org/10.1016/S1098-2140\(02\)00273-4](https://doi.org/10.1016/S1098-2140(02)00273-4)

Larson, A. M., Solis, D., Duchelle, A. E., Atmadja, S., Resosudarmo, I. A. P., Dokken, T., & Komalasari, M. (2018). Gender lessons for climate initiatives: A comparative study of REDD+ impacts on subjective wellbeing. *World Development*, 108, 86–102. <https://doi.org/10.1016/j.worlddev.2018.02.027>

Leahy, C., Winterford, K., Nghiem, T., Kelleher, J., Leong, L., & Willetts, J. (2017). Transforming gender relations through water, sanitation, and hygiene programming and monitoring in Vietnam. *Gender and Development*, 25(2), 283–301. <https://doi.org/10.1080/13552074.2017.1331530>

Ledgerwood, J. (1990). *Changing Khmer conceptions of gender: Women, stories, and the social order*. [Doctoral dissertation, Cornell University].

MacArthur, J., Carrard, N., & Willetts, J. (2021). Exploring gendered change: concepts and trends in gender equality assessments. *Third World Quarterly*, 42(9), 2189–2208. <https://doi.org/10.1080/01436597.2021.1911636>

Mason, K. (2018). Gendered Embodiment. In B. J. Risman, C. M. Froyum, & W. J. Scarborough (Eds.), *Handbook of the Sociology of Gender* (pp. 95–107). Springer International Publishing. https://doi.org/10.1007/978-3-319-76333-0_7

McDowell, L. (1999). *Gender, Identity and Place: Understanding Feminist Geographies*. Minneapolis: University of Minnesota Press.
<https://doi.org/10.15299/local.2015.11.5.181>

Merleau-Ponty, M. (1962). *Phenomenology of perception*. Routledge.

Narain, V. (2014). Shifting the focus from women to gender relations: Assessing the impacts of water supply interventions in the Morni-Shivalik hills of northwest India. *Mountain Research and Development*, 34(3), 208–213.
<https://doi.org/10.1659/MRD-JOURNAL-D-13-00104.1>

National Institute of Statistics [Cambodia]. (2021) *Cambodia Demographic and Health Survey 2021 [Dataset]*. Funded by USAID. Rockville, Maryland: ICF [Distributor].

Patton, M. (2002). *Qualitative Research and Evaluation Methods*. Thousand Oaks: SAGE Publications.

Price, M., Galie, A., Marshall, J., & Agu, N. (2018). Elucidating linkages between women's empowerment in livestock and nutrition: a qualitative study.

Development in Practice, 28(4), 510–524.

<https://doi.org/10.1080/09614524.2018.1451491>

Rao, A., & Kelleher, D. (2005). Is there life after gender mainstreaming? *Gender and Development*, 13(2), 57–69. <https://doi.org/10.1080/13552070512331332287>

Rees, T. (1998). *Mainstreaming equality in the European Union: Education, training and labour market policies*. Routledge.

Rosaldo, M. Z., Lamphere, L., & Bamberger, J. (1974). *Woman, culture, and society*. Stanford University Press.

Roupetz, S., Garbern, S., Michael, S., Bergquist, H., Glaesmer, H., & Bartels, S. A. (2020). Continuum of sexual and gender-based violence risks among Syrian refugee women and girls in Lebanon. *BMC Women's Health*, 20(1), 1–15. <https://doi.org/10.1186/s12905-020-01009-2>

Rowlands, J. (1995). Empowerment examined. *Development in Practice*, 5(2), 101–107. <https://doi.org/10.1080/0961452951000157074>

Rowlands, J. (1997). *Questioning Empowerment: Working with women in Honduras*. Oxfam. <https://doi.org/10.3362/9780855988364.000>

Sen, A. (1990). Development as capability expansion. In K. Griffin; & J. Knight (Eds.), *Human development and the international development strategy for the 1990s*. Macmillan.

UN General Assembly. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*.

UN Women. (2001). *Gender Mainstreaming: Strategy for Promoting Gender Equality*. Office of the Special Advisor on Gender Issues and Advancement of

Women. Working Paper

United Nations. (1995). Beijing Declaration and Platform for Action.

Van der Merwe, S. E., Biggs, R., Preiser, R., Cunningham, C., Snowden, D. J.,

O'Brien, K., Jenal, M., Vosloo, M., Blignaut, S., & Goh, Z. (2019). Making sense of complexity: Using SenseMaker as a research tool. *Systems*, 7(2), 25.

<https://doi.org/10.3390/systems7020025>

Waffi, J. M. (2017). Evaluating Women's Empowerment: Experimenting with a

Creative Participatory Self-evaluation Methodology in Papua New Guinea.

Evaluation Journal of Australasia, 17(3), 40–43.

<https://doi.org/10.1177/1035719X1701700306>

Willan, S., Gibbs, A., Shai, N., Ntini, N., Petersen, I., & Jewkes, R. (2020). Did young

women in South African informal settlements display increased agency after participating in the Stepping Stones and Creating Futures intervention? A

qualitative evaluation. *Social Science and Medicine*, 265, 113302.

<https://doi.org/10.1016/j.socscimed.2020.113302>

Willetts, J., Carrard, N., Crawford, J., Rowland, C., & Halcrow, G. (2013). Working

from strengths to assess changes in gender equality. *Development in Practice*,

23(8), 991–1006. <https://doi.org/10.1080/09614524.2013.840564>

Willetts, J., Halcrow, G., Carrard, N., Rowland, C., & Crawford, J. (2010). Addressing

two critical MDGs together: Gender in water, sanitation and hygiene initiatives.

Pacific Economic Bulletin, 25(1), 211–221.

[https://www.scopus.com/inward/record.uri?eid=2-s2.0-](https://www.scopus.com/inward/record.uri?eid=2-s2.0-77956223968&partnerID=40&md5=51814fe7754249fb0ac3ca3213e5f926)

[77956223968&partnerID=40&md5=51814fe7754249fb0ac3ca3213e5f926](https://www.scopus.com/inward/record.uri?eid=2-s2.0-77956223968&partnerID=40&md5=51814fe7754249fb0ac3ca3213e5f926)

Smyth. (2016). Designing Questions and Questionnaires. In *The SAGE Handbook of Survey Methodology* (pp. 218–235).

World Economic Forum. (2019). *Global Gender Gap Report 2020*. Insight Report. Available at: www.weforum.org (accessed 7 April 2021).

Yin, R. K. (2015). *Qualitative Research from Start to Finish* (Second Edition). Guilford Publications.

<http://ebookcentral.proquest.com/lib/uts/detail.action?docID=2008479>

Figures



Figure 1. Study process and data sources which informed each stage

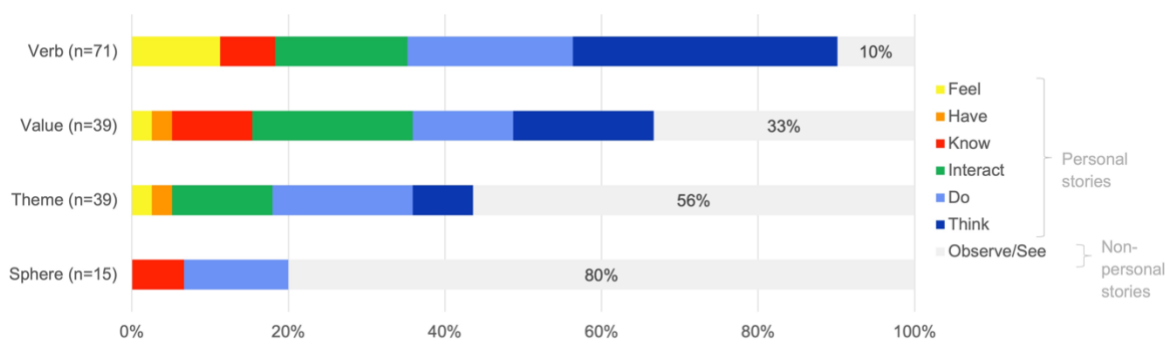


Figure 2. Personal nature of reported changes, by prompt October 2020, SMSU3 Cambodia. Researcher coded. n=164 (excludes 5 blank narratives)

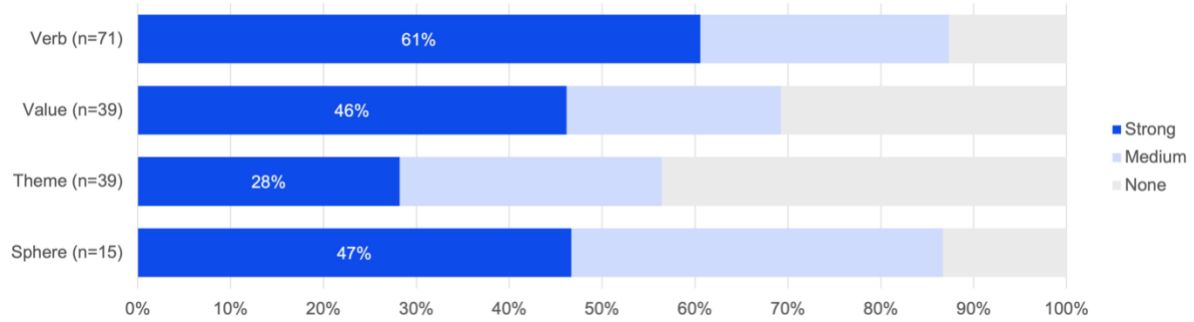


Figure 3. Programmatic contribution to change, by prompt October 2020, SMSU3 Cambodia. Researcher coded. n=164 (excludes 5 blank narratives)

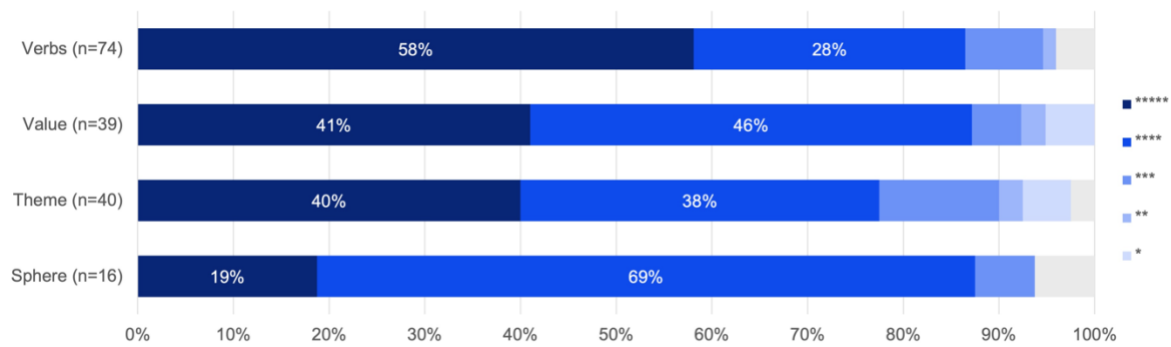


Figure 4. Average story quality, by prompt type, October 2020, SMSU3 Cambodia. Researcher coded. n=169 (includes 5 blank narratives in grey)

Boxes

Box 1. Example of personal and non-personal stories

Example of a non-personal story “When there wasn’t gender [equality] in the workplace, most of the employees were men, and the leader was a man. There weren’t any women who worked as a leader. After the training and the promotion of gender [equality], there are some changes. So, there are more women in the workplace and some leaders are women. These changes are due to the understanding of the equality of gender. It means that women and men have the same knowledge and leadership.” *Sphere-based prompt, Card: change at work*

Example of a personal story “Previously, I thought that women could not be leaders. Now I understand that women can do the same as men. Because I attended the training, I understood that women and men have equal rights” *Verb-based prompt, Card: change in what you think.*

Box 2. Example of program related and non-program-related stories.

No connection to the program “Previously, it was thought that only men could study higher. Nowadays, there is a change, daughters can study like sons. It is due to current promotions.” *Verb-based prompt, Card: change in what you think*

Strong connection to program. “In my family, all the decisions are not only for me, but my wife can also decide.... Previously, all decisions were on me alone. Now all the important decisions in the family must be discussed and agreed smoothly. [This happened] because I participated in training on the gender of men and women.” *Theme-based prompt, Card: change in workloads between women and men*

Box 3. Examples of three levels of story coherence

1* - “There is a change from the work we have done” *Value-based prompt, Negative Iteration*

3*** - “After learning about gender equality, I became more aware of gender equality and participated in exchanging ideas and knowledge with each other, regardless of race or gender.” *Verb-based prompt, Card: change in what you think*

5***** - “My family initially did not want me to work in a community that required travelling long distances because they thought that long-distance travel was unsafe for women and that working in remote areas could have both physical and emotional impacts. He [my father] strictly forbids long-distance travel, which can be detrimental to health, especially community theft and contempt. Now my family has changed a lot because I explained to them the value of women, that women have the right to travel long distances without risk and that my organisation thinks about respecting women's safety. The social reality and behaviour of people is changing the mindset of my elders.” *Theme-based prompt, Card: changes in freedom of movement for women,*

Tables

Table 1. Hear, Create, Assess: Detailed description of the components of the collaborative evaluation process (March – October 2020, SMSU3 Cambodia)

Component		Process
HEAR	Design Workshops	Two in-person workshops (March 2020) considered the cross-cultural challenges in prompt design and tensions between prompt simplicity and specificity. The workshops ultimately helped identify a first round of design criteria which were refined through the CREATE phase.
CREATE	Systematic Review	Systematic literature review of qualitative examples of intervention-based gender equality assessments. The review identified 18 relevant studies from 1,088 potential studies. Studies identified from Web of Science and Scopus in English-language from January 2000 - September 2020. Each article was studied to understand the methods, question phrasing, frequency, and participant types.
	Semantic Review	Detailed syntax review of narrative and retrospective approaches to evaluation, focusing on 23 prompts from published Sensemaker (van der Merwe et al., 2019) and Most Significant Change (Davies and Dart, 2005) techniques.
	Rapid Pilot	Three rounds of rapid piloting (n=28, July – September 2020) to troubleshoot smartphone data collection, visuals, audio options, and question response space (one long answer box, many short answers boxes, and a blend).
	Reflection Interviews	Semi-structured interviews with five rapid pilot participants (September 2020) to refine and simplify the prompt designs. Interviews were live translated from Khmer to English, recorded and transcribed.
ASSESS	Micronarrative Collection	Short story collection from field-based staff members of the project (n=154, October 2020). The micronarrative collection was conducted in small batches (10-15 individuals per batch) in Khmer on smartphones already owned by the staff members. Participants submitted their stories either typing or speaking using an online survey platform in <i>Qualtrics</i> with <i>Phonic</i> (audio recording application) and were encouraged to find private spaces to share their audio recordings. The survey included multiple choice, free response, and audio recording questions.
	Reflection Interviews	Semi-structured interviews with a further 11 respondents to reflect on the prompts and the micronarrative collection process (October 2020).
	Researcher Observations	Observations through daily notes and recorded debrief conversations from the lead researcher and two research assistants (July – October 2020).

Table 2. Design Criteria: Four dimensions of good narratives identified through the evaluation process, March 2020, SMSU3 Cambodia

Dimension	Definition	Assessment Details
Focused: Appropriately focused	In our case, stories are related to gender equality.	Alignment with gender-related concepts and dynamics.
Personal: First-person description of a personal change	Stories depict real life occurrences and are related to the respondent's own personal experiences. This dimension aims to improve story trustworthiness.	First-person description of a personal change. This contrasts with observed changes about another person.
Programmatic: Linked to program influences	In our case, stories are related to the recent gender-mainstreaming interventions for staff. In contrast to stories reflecting more general societal changes related to gender equality.	Inclusion of a description of a project related intervention, such as a training or job role.
Coherent: Sufficiently coherent and detailed	Stories contains sufficient detail and clarity to describe the change. This is not a judgement of the significance of the narrative.	Presence of the five main aspect of storytelling: who, what, where, when, and why. These aim to identify the context, mechanisms, and outcomes of change.

Table 3. Summary of the four selected prompts with regards to objectives, participants, visuals, and prompt wording.
Micronarrative collection October 2020, SMSU3 Cambodia

Name	Value-based	Sphere-based	Theme-based	Verb-based
Objective	Elicit stories of both positive and negative changes.	Elicit stories of change which occur in different locations.	Elicit stories associated with a set of pre-selected outcomes with specific relevance for gender equality.	Elicit stories related to the personal experiences of change.
Participants	n=25 people, 39 stories	n=16, 15 stories	n=25 people, 39 stories	n= 74 people, 71 stories
Visual	<i>emoticons</i>	<i>cards</i>	<i>cards</i>	<i>cards</i>
Pre-prompt	Both this study and SMSU3 are focused on gender equality how experiences and interactions are becoming more equal between women and men.			
Prompt	<p>Thinking about your involvement in the SMSU3 program, what has been the most positive change for you personally, related to gender equality? <i>What were things like before? What are things like now? What specifically changed? Why did this change happen?</i></p> <p>Has anything gotten worse or changed negatively for you personally, related to gender equality? If so, what happened? <i>What were things like before? What are things like now? What specifically changed? Why did this change happen?</i></p>	<p>Thinking about your involvement in the SMSU3 program, where have you seen the most significant change related to gender equality?</p> <p>Within <u>yourself</u> Within <u>your</u> family/home Within <u>your</u> work Within the community</p> <p><i>What specifically changed? What were things like before? What are things like now? Why did this change happen?</i></p>	<p>Thinking about your involvement in the SMSU3 program, pick one of these cards which represents the most significant change for you personally, related to gender equality.</p> <p>Changes in self-awareness of gender inequalities Changes in attitudes towards women leaders Changes in workloads between men and women Changes in decision making between women and men Changes in the level of conflict between men and women Changes in communication between men and women Changes in what is considered normal for men and women Changes in respect for and trust in women Changes in participation of women Changes in confidence for women Changes in freedom of movement for women Changes in personal safety for women Changes in skills and knowledge for women Other (Please describe)</p> <p><i>What specifically changed? What were things like before? What are things like now? Why did this change happen?</i></p>	<p>Thinking about your involvement in the SMSU3 program, pick one of these cards which represents the most significant change for you personally, related to gender equality.</p> <p>Change in what you know Change in what you have Change in what you think Change in how you feel Change in how you speak Change in what you do</p> <p><i>What specifically changed? What were things like before? What are things like now? Why did this change happen?</i></p>

